Kentucky Department for Libraries and Archives

Ensuring the Long-term Accessibility and Usability of Records Stored as Digital Images: Guidelines for State and Local Government Officials

Revised (6-5-2001)

Introduction

Since 1992, state and local agencies have had the explicit right to employ non-erasable optical storage devices to maintain official copies of public records, due to an amendment in the state's public records statute KRS 171.450. Prior to that, however, state and local agencies had begun using imaging technology, and an imaging committee had been formed by the Kentucky Information Resources Management Commission (KIRM) (formerly the Kentucky Information Systems Commission).

The Kentucky Department for Libraries and Archives (KDLA) is authorized in KRS 171.450 to establish "Standards and procedures for recording, managing, and preserving public records and for the reproduction of public records by photographic or microphotographic process." The application of this authority to imaging for the purpose of destruction of the original is based on KRS 171.660, which identifies authorized reproductions as "any records or papers by photographic, non-erasable optical image, or other process which accurately reproduces the original records, which forms a durable medium, and which is performed in accordance with rules and regulations promulgated by the department." Pursuant to this directive, KDLA is issuing these guidelines to help state agencies manage their records in the imaging environment.

Ensuring the long term accessibility and usability of records stored as digital images is largely dependent on how digital imaging systems are designed, implemented, managed, and migrated. A common misperception is that imaged records will be available as long as the physical media used to store the images last. Preservation of media is only one element that ensures long-term accessibility to records. The key to maintaining usable imaged records for long periods is the ability to transport their records, access tools, and required system functionality between hardware platforms, software platforms, and storage media over time. The life of an imaging system is conservatively estimated at about three years, while records retention and access requirements often exceed this short lifecycle. Appropriate policies, management procedures, and technology must be applied from the design of a system until it is redesigned or migrated, to ensure that long-term records are accessible for as long as they are needed. This requires a long-term commitment of resources to preserve the accessibility and usability of digital images.

Legal and Policy Considerations

Specific laws and regulations related to governmental functions may define how records are created, formatted, and maintained. These requirements, as well as legal minimum retention periods established by records retention schedules, should be identified and accounted for when contemplating imaging applications. The Kentucky Revised Statutes (KRS) authorize using of optical imaging as a valid duplication or reproduction media (KRS 171.660) and the records produced in this format are considered a public record. (KRS 61.870, 7.107 & 171.450) Nevertheless, optical images can be challenged in court based on the rules of best evidence and the procedures followed in creating and storing the images. Agencies should be prepared to provide system documentation to courts seeking authentication of legal documents.

Management Considerations

Managerial practices throughout an imaging system's life will have an important impact on the accessibility and use of imaged records. These management practices are especially important for migrating imaged records to new technology environments, which will help ensure long-term retention. Below are the management considerations that should be addressed as systems are planned and developed, as well as those specifically related to system migration.

1. System Planning, Acquisition, and Development

Budget for change

Imaging hardware and software are relatively inexpensive to acquire. The true costs of imaging are centered on training, support, conversion of documents to digital formats, and the continuous system upgrades needed to stay current with the latest technology. Therefore annual budgets should include between 10 to 20% of the cost of the original system for maintenance, support, and upgrade. Because of their expense, imaging systems should also be a part of a reengineering effort to improve agency performance, not just a simple storage media. The scanning process can best be used in conjunction with document, database and forms management systems.

Conduct a Requirements Analysis

The starting point for designing an imaging system is the creation by the agencies of a requirements analysis. The Governor's Office for Technology (GOT) in conjunction with the Kentucky Department for Libraries and Archives have adopted as a standard the model imaging requirements analysis, available at http://www.state.ky.us/ftp/pdf/950805.pdf.

Select a reliable vendor with a good track record

State agencies must select a vendor from state price contract. Local agencies are encouraged to select from the state price contract. Vendor instability is a threat to the long-term viability of imaging systems. Agencies must carefully assess the long-term viability of vendors and the systems they sell when acquiring imaging systems that depend heavily on support from the vendor or manufacturer.

2. System Management

The management of an imaging system is key to the long-term integrity and authenticity of imaged records and their acceptance in legal proceedings and audits. Not all systems merit the same degree of monitoring and control, which should be commensurate with the degree of risk and the benefits to be gained from more effective system management. Special attention should be given to any system that produces records needed in legal proceedings and audits, or one that potentially exposes your organization to a high degree of risk. Because records retention is a critical factor in managing an imaging system, KDLA has issued Policy Memorandum on Optical Storage of Public Records PM 96-1 (updated June 5, 2001) (see Appendix A).

3. Migration

A migration strategy is an essential component in ensuring long-term access to usable imaged records. Such a strategy should guide the movement of records from one generation of technology to another, as well as maintain the functionality for records' use. As mentioned earlier, most imaged records with a retention period of ten or more years will likely have to be migrated at least twice during their lifetime. The possible approaches to migration are expanded by the use of open systems, standard-compliant technology (see below), wise budgeting that accounts for training and technology upgrades, selection of a dependable vendor, and sound management of the system.

Plans for a technology strategy should include a continuum of actions such as:

- ensuring the preservation of imaged records on existing media through careful attention to environmental storage
- maintaining the functionality of existing hardware and software through upgrades of equipment and source code
- transferring the images, index and other related data through successive versions of hardware and software
- migrating optical imaging systems to successive generations of technology, as yet undefined
- Monitoring technology developments and trends, modifying migration plans as needed.

4. Technological Considerations

Migrating applications and records to new technologies in order to retain imaged records over a long period is vital. The technology choices made when systems are developed or upgraded will often determine the ease of and available options for such migrations. If possible, new technology should comply with the following guidelines:

Select an open system solution. An open system solution is one in which the hardware and software components are purchased from different vendors and integrated into a system by a consultant known as a systems integrator. The open systems approach provides a maximum amount of choice to the system developer and end user of the system. Software used in an open system is "portable", which means that it can be moved to a variety of hardware. The software is also "scalable", which means that a system can be sized to handle both small volumes of users and records and expanded to larger volumes. Open systems can therefore be scaled up with limited disruption to operations, including the maintenance of records.

Select standard-compliant system components. System components that are compliant with industry standards and best practices can be more easily upgraded and migrated. State Architectural standards are currently available at http://www.state.ky.us/kirm/arcstand.htm) It is recommend that state and local governments adhere to these standards when acquiring imaging and related technology.

Make controls and system auditing tools available. Systems should be capable of providing audit trails and system security. Effective audit trails can automatically detect who had access to the system, whether staff followed existing procedures, or whether fraud or unauthorized acts occurred or are suspected. Software is available for keystroke monitoring, time and date stamping, virus detection, and other controls that can be built into the design of systems.

Select appropriate storage media and environment. Information and images that are important should be stored on a server (or a mainframe acting as a server) and backed up either on a different computer or on different media. In the past WORM (write once, read many times) technology has been recommended for offline storage of imaged records when long-term retention and legal admissibility is the primary consideration. However, other media may be suitable for such records. If CD-ROM is used as a storage media, it must comply with the ISO 9660 standard, which specifies how a CD-ROM disk stores information. Regardless of media selected, government agencies should:

- never operate drive systems in environments with high levels of airborne particles
- periodically clean optical media to remove dust and other particulates

4. Document Conversion

Document Backfile Conversion Services at KDLA

The Kentucky Department for Libraries and Archives has equipment that allows it to provide comprehensive document-to-image backfile conversion services for Kentucky public agencies. This purchase enables the department to convert paper documents, with planned future purchases to include microfilm conversion. The backfile conversion capacity joins the department's existing program of micrographics services to state and local agencies. With the current equipment, the following services can be performed:

- Scanning of black and white and color documents up to 11' X 17"
- Scanning of black and white and color photographs
- Optical character recognition
- Indexing of documents

In a typical inactive files environment, 10% of inactive files generate 90% of retrieval activity. The cost- effectiveness of converting many files or records that may never need to be accessed should be determined. Backfile conversion can represent significant portions of digital imaging implementation costs. If existing records will be entered into the system, either from paper or microfilm, it must be decided whether to enter the entire record or series, or to purge portions of the record and enter only specific portions.

Document Preparation

In converting paper images to scanned images the following guidelines are important. Those steps that are necessary to make documents "scanner ready" are considered separate from the actual scanning. As in microfilming, proper and thorough document preparation is vital to the success of a scanning program and is the single most important factor in achieving satisfactory throughput rates.

1. At a minimum, documents must be removed from folders, binders or other containers.
2. Documents must be unfolded and stacked in the proper sequence.
3. Each record should be verified for completeness.

4. Missing portions of a record should be located and inserted into its proper place before the other portions of the record are scanned.
5. Automatic feed scanners require the removal of staples, clips, or other fasteners.
6. Torn or damaged documents should be repaired before scanning.
7. The type of record being scanned determines in many cases the scanning order. Transaction or batch documents should be scanned in their sequential order or in the order in which they are received into the agency.
8. Case file applications require that documents that are related to each other be scanned together. This grouping of related documents or sequential ordering of like documents will reduce the amount of media interchange when the documents need to be retrieved from the system.

Regardless of the scanning sequence or the preparation method being used, the entire process should be tested using sample records or documents to verify that no steps have been overlooked. Some documents may present special difficulties in scanning, and may not result in a reasonable image quality unless some sort of enhancement of the document occurs. Documents deliberately not intended for duplication, maps, charts, and documents which the scanner reads the foreground and background as the same color, may not scan successfully.

Scanning Resolution

Scanners are available commercially with real resolution capability between 100 X 100 and 600 x 600 dots per inch (dpi). For *standard text documents*, 200 dpi may be adequate, but this should be considered a minimum resolution. Scanning at 150 or 200 dpi and trying to capture 6 point type won't work well, if at all. Line breakup at 200 dpi can occur on this size or smaller type. If the document has at least 9 point type or larger, 240 dpi or 300 is recommended. Thus the quality of the text depends greatly on the point size and the type of the font being scanned. Also, scanning at the lower dpi will result in the loss of any weak or fine line signatures. A gray scale level of at least 16 bits[EH1]bits and a minimum of 300 DPI is recommended to aid *optical character recognition*.

For capturing detail in drawings or photographs, gray scale and/or a minimum of 300 DPI is necessary. Many scanning systems have prescribed photograph settings which control halftones, gray scale sharpening, brightness and contrast. Testing of prescribed and manual adjustments for the photos is necessary to get the best results. Color photographs may require 24 bit (millions of colors) to accurately capture the image, though 8-bit (256 color) is adequate if images will be used exclusively on the computer screen.

Since higher resolution or use of gray scale can result in larger files, it is important to select the resolution necessary to insure the quality of the record. If the original documents are kept then the resolution requirements may only be that which can be read easily on a display monitor or when printed out. If the originals are destroyed, there are very few options in restoring information that was lost due to low resolution scanning. Some compression algorithms (JPEG particularly) can result in loss of data, particularly if the compression is maximized above a 10:1 ratio. The state standard CCIT Group IV compression is a lossless compression algorithm for text documents if no gray scale is used. If higher resolutions create a network traffic problem, the agency should consider use of a two part or two layer TIFF file to speed retrieval to the screen. The lower resolution image is used for screen display and the higher resolution image is used to store and print a quality image.

The following table shows examples of file sizes at various resolutions and use of gray scale.

File size of a standard 8 1/2 X 11 text document (compressed & uncompressed) at various dpi and use of gray scale

	200 dpi	300 dpi	300 dpi 4 bit gray scale	300 dpi 4 bit gray scale	400 DPI
Uncompressed	500K	1.05 MB	4.08 MB	8.3 MB	2 MB
Compressed CCIT Group IV	50K	105 K	400K	800K	220 KB

Quality Control, Image Inspection and Verification

Quality control refers to the methodology and techniques used to ensure consistency of procedures and output. In a digital imaging operation, documents are typically destroyed after being recorded on the optical medium. Tight quality control procedures are necessary to ensure that the recorded images are of acceptable quality and can be accurately retrieved via the indexing method employed.

Image inspection involves verification of proper and satisfactory scanning. Immediately after scanning, the image is displayed on a video monitor. The operator must confirm that the document is legible, that no corners are folded or parts of the document otherwise obscured, that the document is right-reading, and that the image is of acceptable quality. Until this is accomplished, the document should not be recorded on optical disk. During this inspection and verification sequence, the image should reside on a magnetic buffer (either a tape, cartridge, or internal hard disk). Following inspection, verification, and accurate indexing, the image may be sent to the optical disk storage system.

Image inspection verifies that the scanning procedures and equipment are producing a digitized image of acceptable quality. However, since the recording of these images occurs after the visual inspection, a subsequent inspection step is required to verify acceptable image recording. Following recording of the image on optical disk, the image should be retrieved through the system and verified against the original document to ensure acceptable recording quality. This procedure should be performed daily during the first several weeks after system implementation and periodically thereafter, until consistent recording quality is confirmed.

If images can not be verified as completely readable and accurate after corrective rescanning, then the corrective measures should be documented and the image should be identified as inaccurate and the original hardcopy document should be kept. The location of the original should also be part of the image documentation. If the scanner operator can identify information in the original document that does not appear clearly on the scanned image, then this information should be attached on a separate note with the date and operator's name recorded with the note. This documentation, as with all other documentation, must not be written on the original document, because it will invalidate the authenticity of the image copy as a valid duplicate of the original. Most digital imaging systems permit the attachment of a note to an image as a separate field in the image database. This type of note, which is generally viewable on demand, is the best way to document special quality control information for a particular document. If this feature is not available on the system being used, then a note should be created on paper and scanned as a separate image.

Quality control testing of the scanner, through proper use of the technical targets and user targets described in ANSI/AIIM MS44 1988, should be performed before and after scanning each batch of documents. This will aid in determining whether or not the scanner is maintaining proper calibration during scanning and operating within acceptable parameters. All testing should be documented. The critical factor in determining acceptability of scanned images is the hard copy output of the images. Quality of images cannot be accurately determined by on-screen inspection; only by comparing image output to a quality reference target also output from the system can a determination of acceptability be made.

Training and supervision of operations staff is a key factor in maintaining acceptable image quality. As noted earlier, there are no objective empirical indicators of acceptable image quality for digitally scanned images. An alternative is to categorize documents based upon scanning problems and reach a consensus on how to most effectively capture the "best" image. Ideally, this decision process would involve a team consisting of image system production staff, records managers, and system users and researchers. These evaluations should include visual analysis of workstation display screen images and laser printer output. Retaining a set of representative laser prints for future reference would be a valuable image analysis benchmark tool.

Location of Index Database

Typically, image index databases for optical media systems are stored on a separate magnetic storage media (usually a hard disk drive). For "fail-safe" storage of the database index, index information can also be written to several locations of the optical disk(s) that the index information is derived from.

Defining Indexing Requirements

Indexing parameters are the categories of information by which document images are indexed for retrieval. The parameters should be based on the retrieval needs of current and future users of the system. The selection of index parameters should occur at the time the system is designed. This will allow the system designers to modify system components to accommodate indexing needs. Indexing parameters are the fields that are used to locate and retrieve desired images. An index can also include fields which provide information about a record, such as a summary, notes, or other bibliographic information.

Retrieval only exists for index parameters included in the system. This can be problematic for records requiring long-term storage. Special attention should be taken to insure that current and future access needs are considered when creating a system for records requiring long-term storage. Although index parameters can be added or amended in some programs, this process is costly and difficult to implement.

Index Data Entry

Digital imaging systems use databases to serve as indexes for the digital images stored on optical disks. Generally, a database record is created for each image. This record is broken into fields that correspond to the system's index parameters. Index data entry is the entry of values into these fields. Data entry can be accomplished while scanning is done, immediately after, or at some longer interval. Data entry can be done via key-entry, downloading of values, or auto indexing. Regardless of the data entry method chosen, databases are usually maintained on magnetic media. Access and retrieval speeds for optical disks are significantly slower than magnetic media. The indexes can be written to their associated optical disks for security.

FOR MORE INFORMATION AND ASSISTANCE, CONTACT THE

Technology Analysis and Support Branch
Kentucky Department for Libraries and Archives
300 Coffee Tree Road
Frankfort KY 40602
Telephone: (502) 564-8300

Glossary of Imaging Terms

Audit trails - a procedure or process that documents who used the system, when they used it, what they did while using the system, and what were the results.

Backfile conversion: The process of scanning in, indexing and storing a large backlog of documents on an imaging system. Time consuming. Generally performed by a service bureau.

Compression agorithum - A software or hardware process that “shrinks” images to they occupy less storage space, and can be transmitted faster and easier.

CCIT Group IV compression - an international standard used for image and fax documents.

Eye-readable: Images which can be read by the human eye. In the case of microforms, eye readability requires magnification.

Gray scale - The spectrum or range of shades of black an image has. Scanners' and monitors' gray scales are determined by the number of gray shades or steps they can recognize and reproduce. A scanner which can see a gray scale of 16 will not produce as accurate an image as one that distinguishes a gray scale of 256.

Half-tone: A graphic, usually created from a photograph, in which dots are used to represent the continuous tones that are in the original photograph. (Often expressed in lines per inch)

Image Resolution: The fineness or coarseness of an image as it was digitized, measured as dots per inch (Example 200 x 200 dpi often expressed as 200 dpi). The higher the greater the amount of detail that can be shown.

Optical character recognition (OCR): The ability of a scanner with the proper software to capture, recognize and translate printed alphanumeric characters into machine readable text. OCR uses either "pattern matching" or a feature extraction." With pattern matching, the software has a template of possible characters. A letter is compared to a library of patterns.

Original record: A record prepared in the first instance or any counterpart intended to have the same effect by a person executing or issuing it. If data are stored in a computer or similar device, any printout or other output readable by sight shown to reflect the data accurately is an "original."

Quality: Quality relates to ability of the information technology system to reliably produce and preserve records so that they can be used or recognized by the intended audience.

Scanner - A device that optically senses a human readable image and contains software to convert the image to machine readable code.

Skew: To slant a selected item in a particular direction while scanning. To deskew an image is to correct the skew so that the lines of text run perpendicular to the edge of the page.

Appendix A

Kentucky Department for Libraries and Archives Policy Memorandum on Optical Storage of Public Records

PM 96-1 (Updated June 5, 2001)

Under the terms of KRS 171.660, the Department for Libraries and Archives is directed to establish rules and regulations governing use of optically recorded images and the circumstances under which original, manual, eye-readable copies of imaged documents may be destroyed. As mandated by KRS 171.420, the State Archives and Records Commission is directed to advise the department on matters relating to archives and records management. Because optical storage technologies are evolving so quickly and because the challenge of managing those recordkeeping technologies over time is substantial, the department is issuing the following policy statements. The commission has reviewed and evaluated this memorandum and fully supports its provisions and implementation.

I. Public records which are either, (1) scheduled* as permanent or, (2) whose vital retention status is greater than ten years, must have manual, eye-readable counterparts.

State agencies desiring to maintain optically recorded public records in either of the above-named categories without manual, eye-readable backups may petition the Director, Public Records Division, for relief from this responsibility. The Director will ask the agency to demonstrate in writing the following elements in their imaging procedures:

A. Accuracy and Completeness of Records.

A process of inspection must be in place to confirm that imaged documents are legible and that no corners of the original documents were folded or obscured during scanning. This process should include systematic quality control and audit procedures, as well as operational oversight by staff with detailed knowledge of the process or system used to produce the records.

Resolution and use of gray scale should be appropriate to capture all needed detail within documents. Similarly, scanned images must capture all colors represented in the original documents which are needed to interpret or understand the meaning of the original. The accuracy of the indexing process must also be assured through procedures that visually verify indexes after they have been keyed or created through optical character recognition.

B. Maintenance and Retention of Documentation.

Full and up-to-date process or system documentation should be maintained throughout the life of the records in question. Where processes or the system changes, documentation of processes must be retained until the retention periods have been met. Documentation should conform to standards established for documentation in the Kentucky Department of Information Systems’ published Systems Life Cycle Manual.

C. Audit Trails and Security.

Audit trails documenting who accessed or used the system, when they used it, and what the results of use were must be maintained. Security measures consistent with guidelines and standards provided in Kentucky’s IT Enterprise Architecture and Standards should be adopted and applied.

D. Access to Records.

Records maintained in imaging systems must meet all access requirements defined by Kentucky Revised Statutes or administrative regulations.

E. Backup.

Backup copy of all image files and indexes must be created at intervals determined by the frequency of update and the criticality of the imaging system, as established by the creating agency. Kentucky's IT Enterprise Architecture and Standards and Department of Information Systems guidelines relating to backups and disaster recovery should be used to determine the frequency and means of backing up imaged records. Agencies should contact DIS directly for copies of these guidelines. Off-site backups of permanent and vital records must be maintained.

F. Equipment Maintenance.

An effective maintenance program ensuring that scanners, optical disks, and magnetic storage devices are properly housed and regularly maintained must be in place. Equipment maintenance logs should be kept to document the occurrence of regular maintenance.

G. Training Programs.

Staff employing imaging systems must receive formal training in system use to ensure that standard procedures are routinely followed and that the record or official status of system contents is broadly accepted.

H. Compression Algorithms.

All images should be stored in or convertible to compression formats identified in the IT Enterprise Architecture and Standards. Currently, the standard requires use of, or an ability to convert to, CCITT Group IV.

I. Migration of Data.

Migration paths for all records, including those stored off-line, must be identified and employed to ensure usability of public records throughout their retention period.

If the above conditions are met, the Director, Public Records Division, will provide written authorization for the petitioning agency to maintain records solely in optical format(s). Periodic review of compliance with these standards will be performed by, or on behalf of, Department for Libraries and Archives staff.

II. Transfer of optically stored public records to the State Archives may only be made when agencies have strictly adhered to the IT Enterprise Architecture and Standards and in those cases where agency recordkeeping systems allow relatively simple transference or export of usable records.

Both of these policies should make agencies consider the economies of maintaining records in manual formats for ultimate transfer to the archives or for creating microfilm copies of records early in their life-cycle.

III. Optically stored records whose retention period is less than permanent or without vital records retention status of greater than ten may be maintained without hard copy, eye-readable originals.

Agencies are strongly advised, however, that meeting the criteria described above under policy statement I is critical to the acceptance of the records as valid documentation of agency transactions and is required for the legal admissibility of public records. Agencies should be further advised that records subject to audit are auditable regardless of format and that these criteria will be used by the State Auditor in evaluating the content of imaged records.

*"Schedule" refers to a records schedule formally approved by the State Archives and Records Commission. This schedule provides the agency its only authority to legally destroy public records, in any format. Vital retention status is a designation given certain critical records whose loss would preclude an agency's immediate return to operation or which protect the rights of individuals or groups or document the obligations and transactions of public agencies.

Page: 5

[EH1] Is this supposed to be levels or level?